*DOI:*
10.1177/1073191116658547

*Document Version*
Peer reviewed version

**Cross-validation of two commonly used self-stigma measures, Taiwan versions of the Internalized Stigma Mental Illness scale and Self-Stigma Scale-Short, for people with mental illness**


| Chih-Cheng Chang | Chung-Ying Lin | Petra C. Gronholm | Tsung-Hsien Wu |
|---|---|---|---|
| Chi Mei Medical Center, King's College London, and Chia Nan University of Pharmacy and Science | The Hong Kong Polytechnic University | King's College London | Chi Mei Medical Center |

Author Note

Chih-Cheng Chang, Department of Psychiatry, Chi Mei Medical Center, Tainan, Taiwan; Health Service and Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; Department of Senior Citizen Service Management, College of Recreation and Health Management, Chia Nan University of Pharmacy and Science, Tainan, Taiwan; Chung-Ying Lin, Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hong Kong; Petra C. Gronholm, Health Service and Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; Tsung-Hsien Wu, Department of Psychiatry, Chi Mei Medical Center, Liouying, Tainan, Taiwan.


This research was supported by grant CLFHR10028 from Chi Mei Medical Center, Liouying, Taiwan. Petra C. Gronholm currently receives funding support from the National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed in this paper are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. We thank the staffs of the Departments of Psychiatry, Chi Mei Medical Centers, Tainan and Liouying, for their support with data collection. We also thank Dr. Chi-Wen Chien, who critically reviewed our manuscript and gave us constructive comments.

Correspondence concerning this article should be addressed to Chung-Ying Lin, Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, 11 Yuk Choi Road, Hung Hom, Hong Kong. E-mail: cylin36933@gmail.com

**Running head: Cross-validation of self-stigma measures: ISMI and SSS-S**

**Abstract**

Self-stigma instruments investigate how people with mental illness internalize public stigma. However, information is limited for the psychometric properties of their scores, especially cross-validating scores from different instruments. Thus, we used confirmatory factor analyses (CFAs) and item-response theory (IRT) models to examine the Internalized Stigma Mental Illness scale (ISMI) and the Self-Stigma Scale-Short (SSS-S). Participants with mental illness (n=347) completed both instruments. The CFAs that simultaneously accounted for both the instrument (ISMI and SSS-S) and the trait (Affect, Cognitive, and Behavior concepts) effects outperformed those that accounted only for the instrument effect or only the trait effect. All item scores fit the IRT model, and were fit with ordered, progressing hierarchies in their step difficulties. We conclude that both instruments are feasible for measuring the self-stigma, and that future research can combine the items of both.

*Keywords*: confirmatory factor analysis, severe mental disorders, multitrait-multimethod analysis, item-response theory, self-stigma

## Introduction

Stigma is often described using Goffman's (1963: p. 3) seminal definition of an "attribute that is deeply discrediting", reducing the bearer "from a whole and usual person to a tainted, discounted one". Stigma has also been conceptualized in terms of six critical underlying process components; namely, the co-occurrence of (i) labelling and distinguishing differences; (ii) applying negative stereotypes to those labelled as different; (iii) separating the labelled from the unlabeled ("us" vs. "them"); (iv) emotional reactions amongst both those who do the labelling and those who are labelled; (v) status loss and discrimination experienced by the labelled group, and; (vi) social, political, economic power which enables these processes to unfold (Link & Phelan, 2001; Link, Yang, Phelan & Collins, 2004). Since Goffman's pioneering work there has been a substantial increase in research on this topic for people with mental illness (Bos, Pryor, Reeder, & Stutterheim, 2013) and stigma is recognized as a highly prevalent issue in most countries (Angermeyer & Dietrich, 2006; Jacobsson, Ghanea, & Törnkvist, 2013).

Literature has used multiple means to describe, clarify and classify the broad and multifaceted stigma concept. An overview of this is provided by, for example, Pescosolido and Martin (2015), who outline two perspectives through which stigma can be categorized. The first represents the experiential nature of stigma; that is, whether stigma is perceived, endorsed, anticipated, received, or enacted. The second reflects an action-oriented perspective, in terms of who (or what) gives or receives the stigma. From this perspective, distinctions are made between (i) public stigma (stereotypes, prejudice and discrimination as endorsed by the general population); (ii) structural stigma (prejudice and discrimination through laws, policies, and constitutional practices); (iii) courtesy stigma (stereotypes, prejudice, and discrimination acquired through a connection with a stigmatized group/person); (iv) provider-based stigma (prejudice and discrimination by occupational groups designated to provide assistance to stigmatized groups), and; (v) self-stigma. This last concept of self-stigma is the focus of our present study.

Self-stigma, also called internalized stigma, is defined as a three-step process where a person who belongs to a stigmatized group and is aware of the publicly held stigma regarding this group (1) legitimizes these negative stereotypes and prejudiced attitudes, (2) applies these to themselves, and (3) subsequently experiences negative consequences (Corrigan, Watson & Barr, 2006). Therefore, self-stigmatization is a subjective process embedded in a sociocultural context for people with mental illness, and might result in their negative self-feelings, maladaptive behaviors, or stereotype endorsement (Livingston & Boyd, 2010).

This experience of internalizing publicly held stigma is often discussed as resulting in diminished self-esteem and self-efficacy (Corrigan & Rao, 2012; Corrigan & Watson, 2002; Corrigan et al., 2006) (Corrigan, Larson, Rüsch, 2009). Moreover, people with mental illness are likely to feel embarrassed, guilty, and inferior to those without mental illness (Dinos, Stevens, Serfaty, Weich, & King, 2004; Stevelink, Wu, Voorend, & van Brakel, 2012). A number of further negative consequences have also been examined, including impaired social adaptation, social withdrawal, jeopardized quality of life, reduced well-being, poor mental and physical health, and even suicidality (e.g., Camp, Finlay, & Lyons, 2002; Link, Struening, Neese-Todd, Asmussen, & Phelan, 2001; Livingston & Boyd, 2010; Mak, Poon, Pun, & Cheung, 2007; Wahl, 1999)(Rüsch, Zlati, Black & Thornicroft, 2014). In terms of health-related outcomes specifically, a meta-analysis (Livingston & Boyd, 2010) on 45 articles exploring associations between self-stigma and its correlates reported a positive correlation between the severity of psychiatric symptoms and self-stigma, and a negative correlation between treatment adherence and self-stigma. Certain negative consequences of self-stigma are also evident within healthcare settings. For example, it has been reported that self-stigma can result in unwillingness to undergo appropriate treatment (Corrigan, Larson &

Rüsch, 2009), and self-stigma might impact on the therapeutic effect, which is expected to be better when healthcare professionals take into account the psychiatric symptoms and self-stigma of persons with mental illness (Huang & Lin, 2015; Yanos, Roe, & Lysaker, 2010). Overall, the distress caused by self-stigma is considered attributable to the psychological effect of reflecting on having a socially stigmatized condition (Bos et al., 2013), and its fundamental and inextricable linkage to the concept of discrimination (Peterson, Barnes, & Duncan, 2008).

In view of these numerous negative potential consequences of self-stigma, a better understanding of the nature of self-stigma could help healthcare professionals develop more comprehensive treatment plans for people with mental illness, which aim to not only improve clinical outcomes but also reduce the impact of stigma. Indeed, self-stigma reduction has become a concurrent goal when healthcare professionals mitigate the symptoms or improve the working performance of people with mental illness (Link, Struening, Rahav, Phelan, & Nuttbrock, 1997; Ritsher [Boyd], Otlingma, & Grajales, 2003).

Several instruments are available to measure the degree of self-stigma: the Internalized Stigma of Mental Illness scale (ISMI) (Ritsher [Boyd] et al., 2003), the Self-Stigma Scale-Short (SSS-S) (Mak & Cheung, 2010), and the Self-Stigma of Mental Illness Scale (SSMIS) (Corrigan, Rafacz, & Rüsch, 2011). These instruments have been developed based on somewhat different conceptualizations of self-stigma. The ISMI was developed by Ritsher [Boyd] et al. (2003), based on concepts within two well-established measures on withdrawal and social alienation (the “Different and Ashamed Scale” [Link et al., 2002] and the “Stigma-Withdrawal Scale” [Link et al., 2001]) and ideas from key literature. These materials were developed into a scale with five thematic subscales: *Alienation*, referring to the subjective experience of having a ‘spoiled identity’; *Stereotype Endorsement*, referring to the agreement with different common stereotypes; *Discrimination Experience,* referring to the perception of the way others treat the respondent; *Social Withdrawal,* referring to avoidance of the public; and *Stigma Resistance,* referring to the experience of resisting or being unaffected by self-stigma. The SSS-S (Mak & Cheung (2010) was developed according to cognitive-behavioral theory, conceptualizing three psychological dimensions of self-stigma; after perceiving public stigma, stigmatized individuals may agree with the stereotypes (*Cognitive*); their agreement with these stereotypes may influence their emotions (*Affect*); and finally, the stigmatized individuals are likely to withdraw socially (*Behavior*). The SSMIS adopts the model from Link (Link, 1987; Link & Phelan, 2001), and defines three stages of self-stigma: *Stereotype awareness,* a stigmatized individual's awareness of general negative beliefs; *Stereotype agreement,* a stigmatized individual endorsing the stereotypes that are commonly perceived in the society; and *Self-concurrence,* a stigmatized individual believing that culturally internalized beliefs in fact apply to them (Corrigan et al., 2006).

Although several studies have reported the psychometric properties of the scores of these self-stigma instruments (e.g., Boyd, Emerald, Otilingam, & Peters, 2014; Stevelink et al., 2012; Brohan, Slade, Clement, & Thornicroft, 2010), additional psychometric evaluations are needed for the following reasons. First, albeit the psychometric properties of self-stigma scores have been reported, few studies have focused primarily on examining the psychometric properties of their scores (Chang, Wu, Chen, Wang, & Lin, 2014; Stevelink et al., 2012). Second, to the best of our knowledge, no studies have cross-validated any two or more scores of self-stigma instruments at the same time. Third, most of the studies evaluating the scores of psychometric self-stigma instruments used only classical statistical methods, such as exploratory factor analysis, internal consistency, and test-retest reliability using correlation coefficients (Boyd et al., 2014; Stevelink et al., 2012). However, some advanced statistical methods (e.g., item response theory, which is described in detail later) have been available for more than half a century and, therefore, a cumulative evaluation of the construct validity of those instruments is needed to inform healthcare professionals and researchers (Su, Ng, Yang, & Lin, 2014). Although the advanced statistical methods we propose later have been used many times for other instruments (*cf.* Cheng, Luh, Yang, Su, & Lin, 2015; Jafari, Sharafi, Bagheri, & Shalileh, 2014; Limbers, Newman, & Varni, 2008; Lin, Yang, Lai, Su, & Wang, 2015) the methods have not been used to analyze the ISMI or the SSS-S score. The psychometric properties of these measures thus need to be examined using advanced methods.

There is also further justification for the specific cross-validation of the ISMI and the SSS-S scores specifically due to their similar underlying constructs, evident upon careful comparison of their frameworks. The ISMI was developed with five underlying constructs: Alienation, Stereotype Endorsement, Social Withdrawal, Discrimination Experience, and Stigma Resistance (Ritsher [Boyd] et al., 2003), and the SSS-S with three underlying constructs: Affect, Cognitive, and Behavior (Mak & Cheung, 2010). Although these underlying constructs of the ISMI and the SSS-S do not seem to fully correspond, we suggest that they can be mapped onto each other in the following manner: “Alienation (ISMI)” and “Affect (SSS-S)”, “Stereotype Endorsement (ISMI)” and “Cognitive (SSS-S)”, and “Social Withdrawal (ISMI)” and “Behavioral (SSS-S)”. This is in agreement with Brohan et al. (2010), who likewise proposed that these ISMI domains can be considered reflective of underlying affective, cognitive and behavioral dimensions.

In terms of cross-validating these scores, we propose excluding the “Discrimination Experience” and “Stigma Resistance” constructs within the ISMI, as it has been argued (Brohan et al., 2010; C. C. Chang et al., 2014; Sibitz, Unger, Woppmann, Zidek, & Amering, 2011b) that these are not in line with the aforementioned definition of self-stigma; that is, people legitimizing and internalizing public stigmatizing attitudes, and suffering numerous negative consequences as a result (Corrigan et al., 2006). Namely, “Discrimination

Experience" has been considered embedded within the concept of *experienced* stigma, rather than self-stigma per se. Likewise, "Stigma Resistance" is the *reaction against* stigma, thus also reflecting a process separate to that of self-stigma. By excluding these two constructs from the comparison, it is possible to map the ISMI domains onto the SSS-S concepts (Mak & Cheung, 2010), making it possible to examine their potentially corresponding underlying constructs by cross-validating their scores.

The benefits of concurrently cross-validating two self-stigma instruments include theoretical and empirical advantages. In terms of theoretical benefits, through such efforts researchers can clearly understand the frameworks of the two instruments. That is, whether the two self-stigma instruments share the same constructs, as we propose in this paper, or if the two instruments have distinct underlying constructs. In terms of empirical advantages of cross-validation, if the two instruments share the same constructs, we can further justify whether each construct has insufficient items (e.g., the ISMI items and SSS-S items measure different parts of the same underlying construct), and we may integrate both instruments to measure self-stigma. If the two instruments have different constructs, we may propose a thorough framework using the empirical results from both instruments, and a comprehensive instrument for self-stigma may be developed accordingly.

In terms of advanced statistical approaches that could be used in efforts to examine the psychometric properties of self-stigma scales, classical test theory (CTT) and item response theory (IRT) are two different theories that test psychometric properties of a rating scale. CTT, however, has a major drawback: its summated scores are ordinal and the statistical inference based on ordinal scores may be invalid because of the assumption that raw scores have to be treated as an interval scale (Hobart, Cano, Zajicek, & Thompson, 2007; Jafari, Bagheri, Ayatollahi, & Soltani, 2012). For example, CTT treats a four-point Likert Scale (e.g., *very disagree, disagree, agree, and very agree*) as an interval scale, and scores the four points as one to four. However, the differences between the four points may not be equal. In contrast, IRT uses a logistic equation to estimate respondents' underlying abilities and item difficulties. Although IRT overcomes the CTT problem of using summated ordinal scores, it is often criticized for its complicated model theory, mathematical equations, and output (Ghaemi, 2011). Nevertheless, IRT has the following strengths that allow it to outperform CTT: (1) sample-independent estimation of item parameters, and item-independent estimation of person ability; (2) individually analyzing the validity of each item to determine whether it is redundant or out-of-concept; (3) producing an ordinal-to-interval conversion table that helps clinicians capture the latent traits of respondents (Hambleton, Swaminathan, & Rogers, 1991; Tractenberg, 2010).

In terms of cross-validating instrument scores, options include, for example, concurrent validity, multitrait-multimethod (MTMM) analysis combined with confirmatory factor analysis (CFA), and IRT models. In terms of the appropriateness of these approaches to cross-validate the self-stigma measures considered in this study, concurrent validity, one kind of criterion-related validity, is used to test the correlation coefficients between two related instruments scores that are administrated at the same time (Bollen, 1989). Such an approach might be suited for this study, as both the ISMI and the SSS-S measure the same broad underlying construct of self-stigma, and their scores should as such be highly correlated. MTMM plus CFA is used to test underlying constructs measured with different methods using model comparisons and fit indices (Cheng et al., 2015; Huang & Michael, 2002; Marsh & Grayson, 1995; Su et al., 2014), to examine whether the underlying concepts measured are the same. For example, if a model that tests only certain traits (e.g., Affect, Cognitive, and Behavior) outperforms a model that concurrently tests both traits and methods (e.g., the ISMI and the SSS-S), it may suggest that the two models have redundant items. Contrarily, if a model that tests both traits and methods performs better than one that tests traits only, it may suggest that the two models measure different parts of the same underlying concepts. However, concurrent validity analysis and MTMM plus CFA have to be used under CTT. As the adequacy of CTT is debatable, it has been suggested that a Rasch model, a one-parameter logistic IRT model (Kook & Varni, 2008), could be a better approach for efforts aiming to understand the validity of individual items by estimating interval-level measurement of item difficulty. Specifically, whether the items are out of the underlying construct, or redundant in the underlying construct, can be examined using a Rasch model's assumptions of a hierarchical relationship between the difficulty of the items and the ability of the respondents along the underlying construct (Wright & Linacre, 1994). Therefore, it might be possible to merge the ISMI and SSS-S items if they correspond to the same underlying constructs, and test this using a Rasch model to identify which items are redundant and which are out-of-scope.

Using a two-parameter (2PL) or three-parameter logistic (3PL) IRT model provides more information (namely, the item discrimination) than a Rasch model. However, determining the best approach requires considering the different drawbacks of these three models. 3PL model is inappropriate for health-related patient-reported outcomes because we believe that no one will *guess* his or her health condition. Furthermore, both 2PL and 3PL models depend on unknown parameters for their sufficient statistics, and cannot make an identical unit of difficulties for all items. Therefore, the comparisons between item difficulties and respondent abilities using 2PL and 3PL models become meaningless. Contrarily, the Rasch model generates identical unit (*logit*), and makes the scale as a real interval scale (de Ayala, 2009). In other words, the scores generated from a 2PL model cannot be used as those generated from a Rasch model. In addition, 2PL and 3PL models often need a larger sample size than does the Rasch model (Hulin, Lissak, & Drasgow, 1982; Petscher & Schatschneider, 2012). However, a Rasch model approach contains a strong assumption that all items have a constant discrimination parameter (Sick, 2010), which is a difficult condition to satisfy in reality. Considering all these factors, we decided not to apply a 3PL model, and rather compared Rasch with 2PL models at the beginning of IRT analyses. Afterwards, the results of the best model between Rasch and 2PL models were reported.

In summary, the purpose of this study was to use different approaches and advanced statistical measures —concurrent validity, MTMM plus CFA, and IRT models (Rasch and 2PL) —to cross-validate the scores of two commonly used self-stigma instruments, the ISMI and the SSS-S. Based on the insights generated by this investigation, it is possible to determine whether these instruments should be combined or used separately, and determine which one should be used to use to screen for the level of self-stigma.

## Methods

### Participants and Procedure

The Research and Ethics Review Board of the Chi Mei Medical Center approved this study (IRB number: 10102-L06). All of the participants—psychiatric outpatients, inpatients of psychiatric acute wards, psychiatric daycare patients, and psychiatric patients receiving homecare services (all between 2013 and 2014)—were recruited from Chi Mei Medical Center in Taiwan. The inclusion criteria were (1) twenty years old or above; (2) understand Mandarin Chinese or Taiwanese; and (3) stable psychiatric symptoms during the survey; 350 patients met these criteria. Patients with a diagnosis of dementia, intellectual disability, autism, and organic mental disorder were excluded.

Psychiatrists outlined the purpose of the study to target patients, after which written informed consent was collected from those interested to participate. Subsequently, with the support of research assistants, participants were asked to complete a background information sheet and four self-report questionnaires; described in detail below.

Overall, 350 patients met the inclusion criteria and consented to participate. The study analyses were, however, restricted to data from 347 participants due to missing data on the internalized stigma scale for three participants.

**Instruments**

*Language*

All instruments used in this study were written in Mandarin Chinese using traditional characters; we call these the “Taiwan versions” of these measures. The Mandarin pronunciation, idioms, and Chinese characters used in Taiwan are somewhat different from those used in other areas, such as Hong Kong and mainland China, comparable to the differences between major English dialects. In Taiwan, most people use two spoken languages: Mandarin Chinese and Taiwanese. The Mandarin Chinese spoken in Taiwan is similar to that spoken in mainland China; however, the pronunciation is different from the Cantonese used in Hong Kong. However, spoken Mandarin Chinese and Taiwanese (aka Hoklo and Southern Min) are mutually incomprehensible. In Taiwan, however, they are written using almost the same set of traditional Chinese characters to communicate near-identical meanings. The written traditional Chinese characters used in Taiwan are similar to those used in Hong Kong, but different from the simplified characters used in mainland China. Therefore, speakers of Taiwan’s Mandarin Chinese and Taiwanese in Taiwan can communicate in writing using Taiwan’s traditional Chinese characters. For example, the sentence “I am late for school” is written as “我上學遲到了” in both Taiwan and Hong Kong, but with different pronunciations: “wo3 shang4 xue2 chi2 dao4 le” in Taiwan, and “ngo5 soeng6 hok6 zi6 dou3 liu5” in Hong Kong. However, the pronunciation of the sentence is similar to “wo3 shang4 xue2 chi2 dao4 le” in mainland China, but it is written as “我上学迟到了”.

*Internalized Stigma of Mental Illness Scale*

The original version of the ISMI contains 29 items within five domains (Alienation, Stereotype Endorsement, Discrimination Experience, Social Withdrawal, and Stigma Resistance). Each ISMI item asks about how much the respondents agree with the description measured by a four-point Likert scale (Very disagree, Disagree, Agree, and Very agree). The original version of the ISMI (Ritsher [Boyd] et al., 2003) and the Taiwan version of the ISMI (C.-C. Chang et al., 2014) have acceptable internal consistency ($\alpha$ = .72 to .90 [original version], = .85 to .89 [Taiwan version]) and test-retest reliability (*r* = .68 to .92 [original version], intraclass correlation coefficient [ICC] = .75 to .80 [Taiwan version]) in all domains, except for the Stigma Resistance domain. The mean score of the total ISMI and each domain can be calculated. The total (or domain) score ranges one to four, and a higher score represents a higher level of self-stigma. However, the current study used only three domains of the ISMI—Alienation, Stereotype Endorsement, and Social Withdrawal—with Cronbach's $\alpha$ ranging from .85 to .89. The Resistance Stigma domain is reported to have unsatisfactory psychometric properties (e.g., C.-C. Chang et al., 2014; Ritsher et al., 2003), and both this domain and the Discrimination Experience domain appeared not to capture the self-stigma concept (Brohan et al., 2010; C. C. Chang et al., 2014; Sibitz et al., 2011b).

*Self-Stigma Scale-Short*

The SSS-S is a self-rated questionnaire with nine items and three domains (Affect, Cognition, and Behavior). Each SSS-S item asks about how much the respondents agree with the description using a four-point Likert scale (Very disagree, Disagree, Agree, and Very agree). The original version of the SSS-S (Mak & Cheung, 2010) and the Taiwan version of the SSS-S (Wu, Chang, Chen, & Lin, 2015) have acceptable internal consistency ($\alpha$ =.81 to .84 [original version], = .80 to .91 [Taiwan version]) in all domains and have satisfactory construct validity (comparative fit index [CFI] = .97 [original version], = .99 [Taiwan version]) in a second-order CFA model. The mean score of the total SSS-S and each domain can be calculated. The total (or domain) score ranges one to four, and a higher score on the SSS-S represents a higher level of self-stigma. The internal consistency of the SSS-S in our study ranged from .80 to .91.

*The WHO questionnaire on the Quality of Life, Short Form*

The World Health Organization Quality of Life, Short Form (WHOQOL-BREF) Taiwan version is a 28-item questionnaire with four domains (Physical, Psychological, Social, and Environment). Each WHOQOL-BREF item uses a five-point self-rated Likert scale to assess the respondents' quality of life (QoL). Using participants randomly sampled from 17 hospitals throughout Taiwan, the Taiwan version of the WHOQOL-BREF (Yao, Chung, Yu, & Wang, 2002) has acceptable internal consistency ($\alpha$ = .70 to .91) and test-retest reliability ($r$ = .76 to .80) in all domains. However, its construct validity (CFI = .89) was not promising in a second-order CFA model. In addition, internal consistency ($\alpha$ = .70 to .76) and test-retest reliability (ICC = .81 to .86) of the WHOQOL-BREF Taiwan version were acceptable amongst patients with mental illness, except for the internal consistency of the Social domain ($\alpha$ = .68) (Chang, Wang, Tang, Cheng, & Lin, 2014; Su et al., 2014). Although the internal consistency of the WHOQOL-BREF was somewhat weak in previous studies (Su et al., 2014), in our study this ranged from .77 to .86.

*Depression and Somatic Symptoms Scale*

The Depression and Somatic Symptoms Scale (DSSS) is a 22-item questionnaire with two domains (Depression and Somatic). Using a four-point self-rated Likert scale (Absent, Mild, Moderate, and Severe), respondents indicate how serious given symptoms described by the DSSS items are; higher scores represent a worse condition. Although the DSSS is not as commonly used worldwide as are the Center for Epidemiologic Studies Depression Scale (CES-D) and the Patient Health Questionnaire-9 (PHQ-9), we used the DSSS because it contains more items on somatic complaints than do the CES-D and the PHQ-9. People in East

Asia tend to report their somatic complaints as substitutes for their depressive symptoms, and some may not report their typically depressed mood (Simon, Vonkorff, Piccinelli, Fullerton, & Ormel, 1999). Therefore, using the DSSS seemed culturally appropriate for our study participants. The Taiwan version of the DSSS (Hung, Weng, Su, & Liu, 2006) has acceptable internal consistency ($\alpha$ = .73 to .94) and test-retest reliability (*r* = .88 to .92) in all domains, and satisfactory concurrent validity (*r* = .63 to .86 with the Hamilton Depression Rating Scale). Additionally, a recent study reported that the DSSS is also correlated with the Hospital Anxiety and Depression Scale (*r* = .49 to .79), and the DSSS seems to have better sensitivity in detecting non-full remission and current major depressive disorder than does the Hospital Anxiety and Depression Scale (Hung, Liu, Wang, Yao, & Yang, 2012). In addition, the DSSS was moderately correlated with the mental health subscale of the Short Fort 36 (*r* = −.43; Hung, Wang, & Liu, 2009). The internal consistency of the DSSS in our study ranged from .91 to .93.

**Statistical analysis**

The descriptive analyses and the Pearson correlation analyses were done using SPSS 16.0 for Windows (SPSS Inc., Chicago, IL, USA); confirmatory factor analyses (CFAs), using LISREL 8.8 for Windows (SSI Inc., Lincolnwood, IL, USA); and IRT models, using TAM package in R software (Kiefer, Robitzsch, & Wu, 2016).

*Concurrent validity*

Pearson correlations were used to test the concurrent validity of the ISMI and of the SSS-S total and domain scores. We expected that the ISMI total score would be highly and positively correlated with SSS-S total score; that DSSS domain scores would be moderately and positively correlated with the ISMI and SSS-S total scores; and that the WHOQOL-BREF domain scores would be moderately and negatively correlated with the ISMI and SSS-S total scores.

*Comparisons of CFA models*

Eight CFAs using Maximum Likelihood Estimation were used to examine the construct validity of the ISMI and SSS-S scores. Models 1 and 2 (Figure 1) were the first-order construct of the ISMI (Affect domain has six items, Cognitive domain has seven, and Behavior domain has six) and that of the SSS-S (Affect, Cognitive, and Behavior domains have three items each), respectively. Models 3 and 4 (Figure 2) correlated the self-stigma instruments (ISMI and SSS-S) and the self-stigma traits (Affect, Cognitive, and Behavior), respectively. Therefore, Model 3 assumed that each item is embedded in an instrument construct, and Model 4 assumed that each item is embedded in a self-stigma concept. Models 5 to 8 (Figure 3 and Figure 4) were used to simultaneously examine the construct validity of the two self-stigma instruments scores: Model 5 assumed that each item was embedded in one general self-stigma trait and one self-stigma instrument concept (ISMI and SSS-S); Model 6 assumed that each item was embedded in one specific self-stigma trait (Affect, Cognitive, or Behavior) and in one instrument concept (ISMI or SSS-S); and Models 7 and 8 assumed that each item was embedded in one specific self-stigma trait and in one instrument concept. In addition, Models 6 to 8 assumed that all specific self-stigma traits were correlated. Models 5 and 8 assumed that the two instrument concepts were correlated, while Model 6 assumed that they were uncorrelated. Therefore, comparing Models 3-8 helps us understand whether self-stigma traits are differently measured by the two self-stigma instruments.

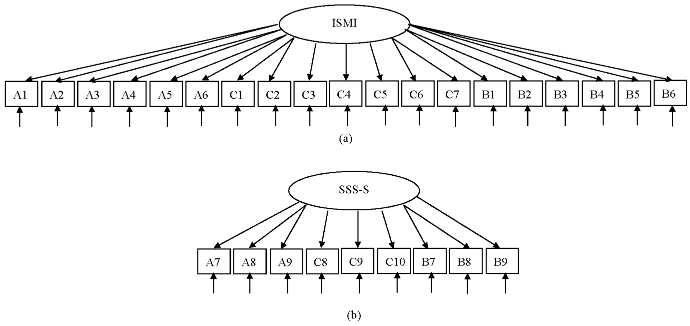


Figure 1

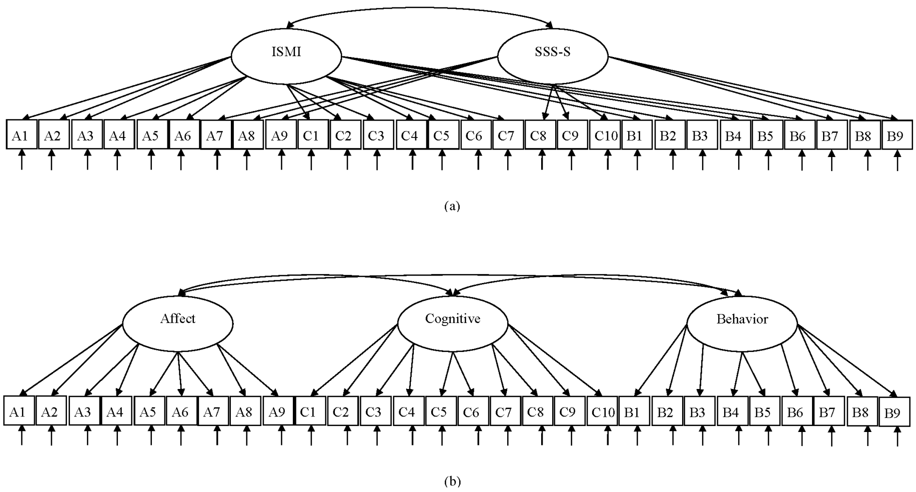


Figure 2

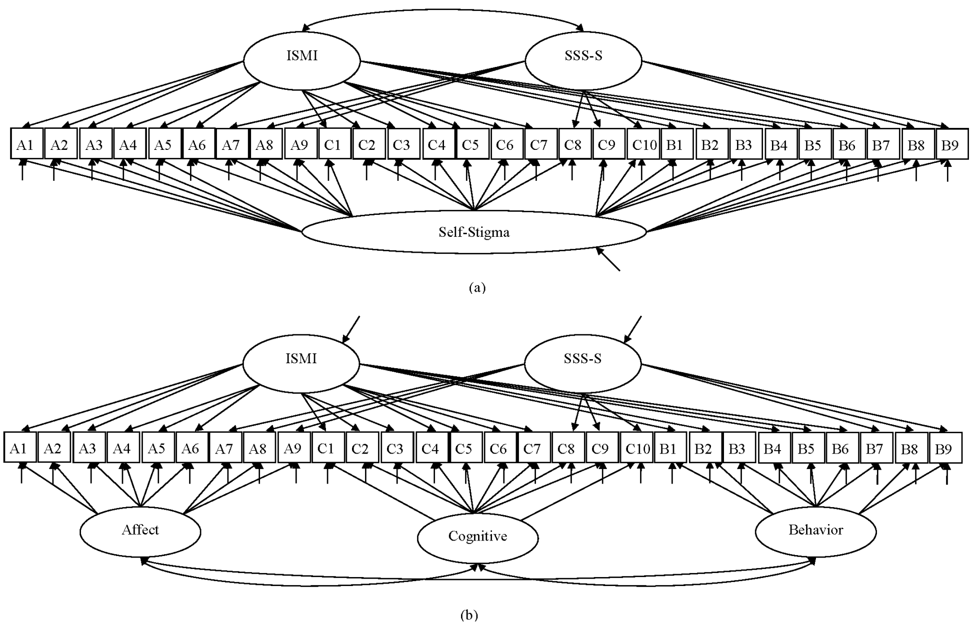


Figure 3

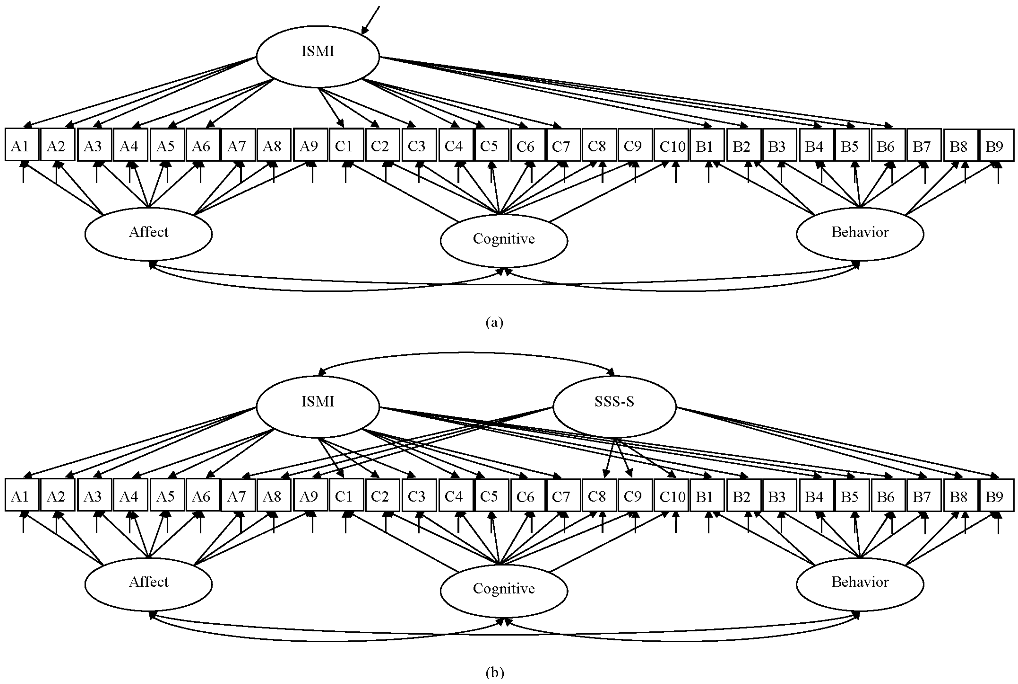


Figure 4

The tested CFA models were examined using several indices: the comparative fit index (CFI), incremental fit index (IFI), and Tucker-Lewis index (TLI) all being greater than or equal to .90 (Garcia-Barrera, Kamphaus, & Bandalos, 2011; Su et al., 2014); and the root mean square error of approximation (RMSEA) and standardized root-mean-square residual (SRMR) both being less than or equal to .08 (Hu & Bentler, 1999; Lin et al., 2013) suggested that they were data-model fit acceptable. The comparisons between the CFA models rely on $\chi^2$ difference tests (Bollen, 1989), where a $\chi^2$ difference value and a *df* difference value of two compared models are computed, and if the two compared models are significantly different, the $\chi^2$ difference value will have a *p*-value less than .05 based on the *df* difference (e.g., when comparing Models 1 and 2, $\chi^2$ difference [$\Delta\chi^2$] = Model 1 $\chi^2$ – Model 2 $\chi^2$; *df* difference [$\Delta df$] = Model 1 *df* – Model 2 *df*, and using $\Delta\chi^2$ and $\Delta df$ in terms of *p*-value to determine the significant difference between Models 1 and 2).

*Tests of unidimensionality and the comparisons of Rasch and 2PL models*

We first examined the unidimensionality assumption of the three self-stigma domains using one-factor CFA, before comparing Rasch and 2PL models. Two indicators of goodness-of-fit index (GFI) and root-mean-square residual (RMSR) were used to determine the unidimensionality. Specifically, values of GFI over .90 and those of RMSR under .10 suggest essential unidimensionality (Samuel, Simms, Clark, Livesley, & Widiger, 2010). After confirming the unidimensionality, we used Akaike information criterion (AIC), Bayesian information criterion (BIC), and $\chi^2$ difference test to decide the best model between Rasch and 2PL models. The AIC, BIC, and $\chi^2$ indicate better fit in lower values, and a significant $\chi^2$ difference indicates a statistical differece between Rasch and 2PL models.

*Analyses of Rasch or 2PL models*

Several Rasch partial credit models or 2PL models were used to test whether each item fit the underlying constructs. That is, the three hypothesized self-stigma domains (Affect, Cognitive, and Behavior) were separately analyzed using appropriate IRT models. The item fit was examined using both the information-weighted fit statistic (infit) mean square (MnSq) and the outlier-sensitive fit statistic (outfit) MnSq. For clinical measures, a reasonable range for both infit MnSq and outfit MnSq is 0.6 to 1.4 (Khan, Chien, & Brauer, 2013; Chien & Bond, 2009; Wright & Linacre, 1994), where an item fit less than 0.6 suggests a redundant item; greater than 1.4 suggests a misfitting item. IRT analyses were rerun after removing redundant and misfitting items to ensure the acceptable fit of the remaining items. Moreover, IRT analyses presented item difficulties (including average difficulties and step difficulties) of each item with log-odd units (*logits*).

**Justification of sample size**

Because both CFA and IRT models are complicated statistical analyses, and a sample size equal to or greater than a stipulated minimum (say, over 250) is considered necessary to reach a stable estimate, we justified our sample size as sufficient. For the CFA part, Anderson and Gerging (1984) agree with MacCallum, Widaman, Zhang, and Hong (1999) that a sample size of 100 is sufficient for convergence in a CFA model with three or more indicators per factor. Kline (1998) says that a sample between 100 and 200 is medium-sized. Although it seems that 100 is be sufficient for CFA, some empirical researchers (e.g., Fang, Lin, Chen, & Lin, 2015; Su et al., 2014) feel that a sample of at least 200 is necessary. For the IRT part, Reeve & Fayers (2005) agree with Embretson and Reise (2000) that an IRT model with a five-point Likert-scale can produce a reasonable estimate with 250 respondents, but that at least 500 are required for accurate parameter estimates. Because we used a four-point

Likert-scale, which is less complicated than a five-point Likert-scale, we can justify a smaller sample size for our study than if we had used a five-point Likert-scale. In addition, Linacre (2002: p.88) provides another rule of thumb to calculate sufficient sample sizes for Rasch partial credit models: "when there are m+1 categories and so m steps…. This would make the minimum sample size for stability at least 25 × (m+1) subjects, and perhaps as many as 100 × (m+1) subjects if use of categories is not uniform across the rating scale", where the "m" denotes the thresholds in the response. That is, a four-point Likert type scale has the "m" as "3" because there are three thresholds (from 1 to 2, 2 to 3, and 3 to 4). If we apply the above suggestion to the ISMI and SSS-S items (each item has three steps and four categories), the range of a sufficient sample size is between 100 and 400.

## Results

### Demographics and self-stigma scores

The mean age (± standard deviation [SD]) of the participants was 43.76 (± 11.27) years, and the onset of their mental illness was 31.88 (± 11.81) years. There were more female ($n$ = 199: 57.3%) than male participants ($n$ = 148); 28.8% of the participants had a college degree or higher ($n$ = 101); 138 were married and 156 were single. Almost half ($n$ = 161: 46.4%) had been diagnosed with schizophrenia, more than a quarter with depressive disorder ($n$ = 98: 28.2%), and an eighth with bipolar disorder ($n$ = 43: 12.4%) (Table 1).

(Insert Table 1 here)

Participants reported significantly higher self-stigma scores for the ISMI than for the SSS-S (differences = −0.09 [total score], −0.12 [Affect domain], −0.20 [Cognitive domain]; $ps < .001$), except for the Behavior domain (difference = 0.05, $p = .06$) (Table 1).

### Concurrent validity

The ISMI total score was highly correlated with the SSS-S total score in observed scores ($r = .859$). The intercorrelations between the ISMI and SSS-S domain scores were high ($r$

= .836 to .852 [ISMI], .815 to .862 [SSS-S], and .752 to .821 [ISMI and SSS-S]). In addition, both ISMI and SSS-S domain scores and total scores were moderately correlated with the WHOQOL-BREF domain scores (*r* = −.474 to −.641 [ISMI], −.447 to −.583 [SSS-S]) and the DSSS domain scores (*r* = .386 to .559 [ISMI], .335 to .457 [SSS-S]) (Table 2).

(Insert Table 2 here)

**Comparisons of CFA models**

All of the fit indices in Models 1 and 2, which respectively represent the constructs of the ISMI and of the SSS-S scores, were acceptable, except for the slightly high RMSEA (.089) in Model 2. In addition, all of the fit indices in Models 3 and 4, which respectively tested the method effects (ISMI and SSS-S) only and the traits (Affect, Cognitive, and Behavior) only, were acceptable, except for the slightly high RMSEA (.100).

All of the fit indices in Models 5 to 8 were satisfactory (Table 3), and showed a significant improvement over Model 3 ($\Delta\chi^2$ = 199.23, 244.49, 49, 69.84, and 245.10; $\Delta df$ = 28, 30, 21, and 31; *ps* < .001), which suggested that self-stigma traits existed, as well as a significant improvement over Model 4 ($\Delta\chi^2$ = 441.25, 486.51, 311.83, and 487.12; $\Delta df$ = 26, 28, 19, and 29; *ps* < .001), which suggested the method effects existing. The comparisons between Models 5, 6, and 8 showed that Models 6 and 8 had similar fit indices, and no significant differences ($\Delta\chi^2$ = 0.61; $\Delta df$ = 1; $p$ = .43), but both Models 6 and 8 outperformed Model 5 ($\Delta\chi^2$ = 45.26 and 45.87; $\Delta df$ = 2 and 3; *ps* < .001; Table 4).

(Insert Tables 3 and 4 here)

**Tests of unidimensionality and the comparisons of Rasch and 2PL models**

All the three self-stigma domains had acceptable GFI (Affect=.909; Cognitive=.908; Behavior=.911) and RMSR (Affect=.033; Cognitive=.035; Behavior=.030) values, and fulfilled the assumption of unidimensionality for IRT models. Therefore, we continually compared the Rasch and 2PL models. The values of AIC (6202.221 for Affect, 6945.251 for Cognitive, and 5837.274 for Behavior), BIC (6344.646 for Affect, 7103.074 for Cognitive, and 5979.699 for Behavior), and $\chi^2$ (6128.221 for Affect, 6863.251 for Cognitive, and 5763.274 for Behavior) in the Rasch models were larger than those of AIC (6138.736 for Affect, 6867.861 for Cognitive, and 5716.130 for Behavior), BIC (6311.956 for Affect, 7060.327 for Cognitive, and 5889.349 for Behavior), $\chi^2$ (6048.736 for Affect, 6767.861 for Cognitive, and 5626.130 for Behavior) in the 2PL models. Moreover, the $\chi^2$ difference tests showed that all 2PL models significantly outperformed all the Rasch models (all *p*s < .001). Therefore, we reported the results based on 2PL models.

**Results of 2PL models**

Separate 2PL model analyses of the three domains, by combining relevant ISMI and SSS-S items, revealed that the scores of all items of the Affect, Cognitive, and Behavior domains showed acceptable goodness-of-fit with both infit MnSq and outfit MnSq values (Table 5). Moreover, all items had ordered progressing hierarchies in their corresponding difficulties; that is, category 1 (Very disagree) had the smallest difficulty, followed by categories 2 (Disagree), 3 (Agree), and 4 (Very agree; Table 5).

(Insert Table 5 here)

**Discussion**

To the best of our knowledge, this is the first study that has simultaneously cross-validated two commonly used self-stigma instrument scores (three domains of the ISMI and the SSS-S) by using a combination of CTT and advanced statistical methods. The purpose of this study was to use a range of approaches and advanced statistical methods to cross-validate the ISMI and the SSS-S, and use these insights to determine whether these instruments should be combined or used separately, and how appropriate they are for screening individuals' self-stigma. Our findings of good validity for both instruments agree with previous studies focused on individual instruments (C.-C. Chang et al., 2014; Mak & Cheung, 2010; Ritsher [Boyd] et al., 2003; Wu et al., 2014). Thus, our results suggest that both the ISMI and the SSS-S are effective and feasible for assessing the self-stigma of people with mental illness. It should, however, be noted that we only tested three of five domains in the ISMI, as these were considered to best capture self-stigma.

Our findings support the theoretical structure that corresponds with cognitive-behavioral theory (Barlow et al., 2004; Mak & Cheung, 2010), which allows the ISMI and SSS-S items to be empirically categorized into the dimensions of Affect, Cognitive, and Behavior, and they support the notion that these three domains are related to self-stigma. In other words, we justified that ISMI shares the same framework with the SSS-S. The other important finding was that no items on the ISMI and the SSS-S are redundant in their underlying traits, which indicates that researchers or healthcare professionals can profitably use both instruments simultaneously. That is, using both instruments may facilitate a thorough assessment of an individual's level of self-stigma. Alternatively, either instrument can be selected based on suitability; for example, using the SSS-S for screening in busy practices or using the ISMI for obtaining maximum information.

Because the ISMI and the SSS-S are designed to measure self-stigma in people with mental illness (Mak & Cheung, 2010; Ritscher [Boyd] et al., 2003), their scores are highly correlated, as expected. In addition, our finding of a moderate correlation between their scores and the WHOQOL-BREF and DSSS scores agrees with the findings of others (e.g., Livingston & Boyd, 2010; Ritsher [Boyd] & Phelan, 2004; Sibitz et al., 2011a) that people with mental illness who are highly self-stigmatized tend to have a worse QoL and more depressive moods than do those who are only mildly self-stigmatized. Therefore, the concurrent validity of their scores was confirmed.

The strength of this study was using two different kinds of advanced statistical methods – MTMM and IRT – to test the psychometric properties of two commonly used self-stigma instruments: no other studies have used MTMM or IRT models to evaluate the psychometric properties for any existing self-stigma instrument. Using MTMM plus CFA models, additional evidence of construct validity can be provided (Huang & Michael, 2002; Marsh & Grayson, 1995), and quantitative descriptions of the method and trait effects can be portrayed for self-stigma instrument scores. Moreover, our MTMM results supported the underlying concepts of Affect, Cognitive, and Behavior based on cognitive-behavioral theory (Barlow et al., 2004; Beck, 1993). Although two of our CFA models (Models 6 and 8) share similar fit indices, we judged Model 8, which has correlated self-stigma traits (Cognition, Affect, and Behavior) and correlated methods (ISMI and SSS-S), to be theoretically preferable. Namely, because the ISMI and the SSS-S were both self-reported, their latent constructs should be correlated based on the common-method variance. Mak and Cheung (2010) also claim that the three underlying concepts related to self-stigma may have a causal relationship. For example, this relationship starts from self-stigmatizing cognitions (e.g., stigmatized people perceive their weakness as incompetence and thus endorse and internalize negative beliefs about themselves); this engenders a self-stigmatizing affect (e.g., they feel shame, despondency, embarrassment, or anger, responses caused by their self-deprecating perceptions), and ends up with self-stigmatizing behaviors (e.g., withdrawal from social interaction) (Lin, Chang, Wu, & Wang, 2016). Because the present study is cross-sectional, it cannot provide any direct evidence for the causal relationship; therefore, more investigation is necessary.

Using IRT models provides two benefits. First, item characteristics and item validity can be investigated, and appropriate items with goodness-of-fit can be detected. Moreover, the interval-level measurement of the stigma level of each participant, and the interval-level measurement of item difficulties are established. Second, computerized-adaptive testing (CAT) may also be used for a self-stigma evaluation that is more efficient. The step difficulties of all items were between −5.80 and 6.42 *logits*, and the items we analyzed suited respondents with a stigma level between −6 and 7 *logits*, a range appropriate for over 99% of the population. In addition, using all items (including the ISMI and the SSS-S items) to measure self-stigma seems adequate. However, asking people with mental illness to complete all of the test items might require a substantial amount of time, especially when assessing stigma that asks for psychological self-perception of negative self-feelings, maladaptive behaviors, or stereotyped endorsement. Therefore, a future trend could be to use CAT with IRT models to substantially reduce the measure administration time, according to Ware et al. (2003) who conducted a practical CAT program and found that it reduced the number of tested items by more than 90%, without compromising validity. CAT asks the respondents the most suitable items based on the item characteristics analyzed using the IRT model and the respondents' previous answers. Therefore, the stigma level of the respondents can be precisely estimated using a CAT program. Using our results of Affect domain as an example, the first question could be A9 ("I feel like I cannot do anything about my mental illness status"). If the respondent answers 3 (Agree); we know that his/her latent stigma level may be between 1.88 and 2.41. Then, the second question could be A1 ("I feel out of place in the world because I have a mental illness"). If the respondent answers 3 (Agree) again, we can estimate that the latent stigma level is between 2.10 and 2.41. A third question of A7 ("I feel that others would know that I have a mental illness") may be continuously used. If the respondent answers 4 (Very agree); then, we can estimate the latent stigma level being 2.31 and 2.41 for the respondent. However, aforementioned is just a simple example, and establishing a CAT program is much more complicated (*cf*. Rudick, Yam, & Simms, 2013).

Our results should, however, be interpreted with caution, because our sample size was relatively small (albeit within an adequate range) and the generalizability of our participant group might be limited. Specifically, the item parameters retrieved from our study were based on 347 people with different kinds of psychiatric diseases. Therefore, the item parameters may not be stable due to the heterogeneity of our sample, and the parameters are likely to change if the sample size was increased. Nevertheless, we believe our findings shed some light for clinicians and researchers to know the benefits of using IRT and CAT program, and collecting information with a large sample size may overcome the aforementioned limitation.

Other limitations are listed as follows. First, because the current study only analyzed three of the five domains of the ISMI, our assessment of the construct validity of the ISMI might be untenable. We do, however, feel it was justified to focus on three domains only (namely, including Alienation, Stereotype Endorsement, and Social Withdrawal, and excluding Discrimination Experience and Stigma Resistance), as (1) the other two domains are suggested to be outside the self-stigma concept (Brohan et al., 2010; Sibitz et al., 2011b); (2) the analyzed domains of the ISMI correspond to the concepts of Affect, Cognitive, and Behavior, based on cognitive-behavioral theory (Barlow, Allen, & Choate, 2004); and (3) our CFA model with three ISMI domains (i.e., Model 1) showed good construct validity. Nevertheless, we acknowledge that the information captured by the entire ISMI may be of interests for clinicians and researchers. Therefore, we conducted an additional set of analyses using the entire ISMI, following the same analytic strategies as in the main study. These additional results (reported in detail in Appendix A) were comparable to the findings of our main analyses. However, the CFA results using the entire ISMI were worse than those based on three domains only. This could be argued to correspond with our justification that the Discrimination Experience and Stigma Resistance domains may be out of the self-stigma concept.

Second, the 2PL IRT model that can freely estimate the discrimination parameter for each item (Yen & Fitzpatrick, 2006) was used in the analyses. However, a sample (n > 500) larger than ours (n = 347) is recommended for the 2PL IRT model (Hulin et al., 1982; Petscher & Schatschneider, 2012). Therefore, future studies should aim to gather a larger sample and attempt to replicate our analyses using the 2PL IRT model for validating ISMI and SSS-S scores.

In summary, our results indicate that both the ISMI and the SSS-S are appropriate for measuring the self-stigma of people with mental illness across three domains of cognition, affect, and behavior that correspond to cognitive-behavioral theory. The construct validity of the scores of these two self-stigma instruments was supported, and IRT analysis showed that almost all items from the two instruments had acceptable goodness-of-fit, in a model-expected hierarchical order from easy to difficult. Future research might want to combine the items of the ISMI and the SSS-S, and use CAT to efficiently evaluate the self-stigma of people with mental illness.

**Figure Legends**

**Figure 1.** (a) Model 1: construct of the Internalized Stigma Mental Illness scale (ISMI). (b) Model 2: construct of the Self-Stigma Scale-Short (SSS-S). A = Affect; C = Cognitive; B = Behavior.

**Figure 2.** (a) Model 3: correlated methods of the Internalized Stigma Mental Illness scale (ISMI) and the Self-Stigma Scale-Short (SSS-S). (b) Model 4: correlated traits of Affect, Cognitive, and Behavior domains of self-stigma. Abbreviations: see legend for Figure 1.

**Figure 3.** (a) Model 5: one general trait (self-stigma) and correlated methods. (b) Model 6: correlated traits and uncorrelated methods. Abbreviations: see legend for Figure 1.

**Figure 4.** (a) Model 7: correlated traits and correlated methods minus 1. (b) Model 8: correlated traits and correlated methods. Abbreviations: see legend for Figure 1.

**REFERENCES**

Anderson, J. C. & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*(2), 155-173.

Angermeyer, M. C., & Dietrich, S. (2006). Public beliefs about and attitudes towards people with mental illness: a review of population studies. *Acta Psychiatrica Scandinavica, 113,* 163-179.

Barlow, D. H., Allen, L. B., & Choate, M. L. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy, 35*, 205-230.

Beck, A. T. (1993). *Cognitive therapy and the emotional disorders.* New York: Penguin.

Bollen, K. A. (1989). Measurement models: the relation between latent and observed variables. In K. A. Bollen (Ed.), *Structural equations with latent variables* (pp. 179-225). New York: John Wiley & Sons.

Bos, A. E., Pryor, J. B., Reeder, G. D., & Stutterheim, S. E. (2013). Stigma: advances in theory and research. *Basic and Applied Social Psychology, 35*(1), 1-9.

Boyd, J. E., Emerald, P. A., Otilingam, P. G., Peters, T. (2014). Internalized stigma of mental illness (ISMI) scale: a multinational review. *Comprehensive Psychiatry, 55*, 221-231.

Brohan, E., Slade, M., Clement, S., & Thornicroft, G. (2010). Experience of mental illness stigma, prejudice and discrimination: a review of measures. *BMC Health Service Research, 10*, 80.

Camp, D. L., Finlay, W. M., & Lyons, E. (2002). Is low self-esteem an inevitable consequence of stigma? An example from women with chronic mental health problems. *Social Science & Medicine, 55,* 823-834.

Chang, C.-C., Su, J.-A., Tsai, C.-S., Yen, C.-F., Liu, J.-H., & Lin, C.-Y. (2015). Rash analysis suggested three unidimensional domains for Affiliate Stigma Scale: additional psychometric evaluation. *Journal of Clinical Epidemiology, 68*(6), 674-683.

Chang, C.-C., Wu, T.-H., Chen, C.-Y., Wang, J.-D., & Lin, C.-Y. (2014). Psychometric evaluation of the Internalized Stigma of Mental Illness Scale for patients with mental illnesses: measurement invariance across time. *PLoS One, 9*(6), e98786.

Chang, K.-C., Wang, J.-D., Tang, H.-P., Cheng, C.-M., & Lin, C.-Y. (2014). Psychometric evaluation using Rasch analysis of the WHOQOL-BREF in heroin-dependent people undergoing methadone maintenance treatment: further item validation. *Health and Quality of Life Outcomes*, *12*, 148.

Cheng, C.-P., Luh, W.-M., Yang, A.-L., Su, C.-T., & Lin, C.-Y. (2015). Agreement of children and parents scores on Chinese version of Pediatric Quality of Life Inventory 4.0: further psychometric development. *Applied Research in Quality of Life*. First online: 24 April 2015.

Chien, C.-W., & Bond, T. G. (2009). Measurement properties of fine motor scale of Peabody developmental motor scales-second edition: a Rasch analysis. *American Journal of Physical Medicine and Rehabilitation, 88*(5), 376-386.

Corrigan, P. W., & Rao, D. (2012). On the self-stigma of mental illness: stages, disclosure, and strategies for change. *Canadian Journal of Psychiatry, 57*(8), 464-469.

Corrigan, P. W., & Watson, A. C. (2002). The paradox of self-stigma and mental illness. *Clinical Psychology: Science & Practice, 9*, 35-53.

Corrigan, P. W., Larson, J. E., & Rüsch, N. (2009). Self-stigma and the “why try” effect: impact on life goals and evidence-based practices. *World Psychiatry, 8*, 75-81.

Corrigan, P. W., Rafacz, J., & Rüsch, N. (2011). Examining a progressive model of self-stigma and its impact on people with serious mental illness. *Psychiatry Research, 189*, 339-343.

Corrigan, P. W., Watson, A. C., & Barr, L. (2006). The self-stigma of mental illness:

implications for self-esteem and self-efficacy. *Journal of Social & Clinical Psychology, 25*, 875-884.

de Ayala, R. J. (2009). *The theory and practice of item response theory.* New York, NY: The Guilford Press.

Dinos, S., Stevens, S., Serfaty, M., Weich, S., & King, M. (2004). Stigma: the feelings and experiences of 46 people with mental illness. Qualitative study. *British Journal of Psychiatry, 184*, 176-181.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Fang, S.-Y., Lin, Y.-C., Chen, T.-C., & Lin, C.-Y. (2015). Impact of marital coping on the relationship between body image and sexuality among breast cancer survivors. *Supportive Care in Cancer, 23*(9), 2551-2559.

Garcia-Barrera, M. A., Kamphaus, R. W., & Bandalos, D. (2011). Theoretical and statistical derivation of a screener for the behavioral assessment of executive functions in children. *Psychological Assessment, 23*, 64-79.

Ghaemi, H. (2011). Is Rasch model without drawback? A reanalysis of Rasch model limitations. *Modern Journal of Language Teaching Methods, 1*(2), 31-38.

Goffman, E. (1963). *Stigma: notes on the management of spoiled identity.* Englewood Cliffs NJ: Prentice Hall.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hobart, J., Cano, S., Zajicek, J., & Thompson, A. (2007). Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurology, 6*, 1094-1105.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Huang, C.-J., & Michael, W. B. (2002). Multitrait-multimethod analyses of scores on a Chinese version of the dimensions of self-concept scale. *Educational & Psychological Measurement*, *62*, 355-372.

Huang, W.-Y., & Lin, C.-Y. (2015). The relationship between self-stigma and quality of life among people with mental illness who participated in a community program. *Journal of Nature and Science*, *1*(7), e135.

Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: a Monte Carlo study. *Applied Psychological Measurement, 6*, 249-260.

Hung, C.-I., Liu, C.-Y., Wang, S.-J., Yao, Y.-C., & Yang, C.-H. (2012). The cut-off points of the Depression and Somatic Symptoms Scale and the Hospital Anxiety and Depression Scale in detecting non-full remission and a current major depressive episode. *International Journal of Psychiatry in Clinical Practice, 16*, 33-40.

Hung, C.-I., Wang, S.-J., & Liu, C.-Y. (2009). Validation of the Depression and Somatic Symptoms Scale by comparison with the Short Form 36 Scale among psychiatric outpatients with major depressive disorder. *Depression and Anxiety, 26*, 583-591.

Hung, C.-I., Weng, L.-J., Su, Y.-J., & Liu, C.-Y. (2006). Depression and somatic symptoms scale: a new scale with both depression and somatic symptoms emphasized. *Psychiatry and Clinical Neurosciences, 60*, 700-708.

Jacobsson, L., Ghanea, H., & Törnkvist, B. (2013). Internalized stigma of mental illness in Sweden and Iran—a comparative study. *Open Journal of Psychiatry, 3*, 370-374.

Jafari, P., Bagheri, Z., Ayatollahi, S. M., & Sltani, Z. (2012). Using Rasch rating scale model to reassess the psychometric properties of the Persian version of the PedsQL$^{TM}$ 4.0

Generic Core Scales in school children. *Health & Quality of Life Outcomes, 10*, 27.

Jafari, P., Sharafi, Z., Bagheri, Z., & Shalileh, S. (2014). Measurement equivalence of the KINDL questionnaire across child self-reports and parent proxy-reports: a comparison between item response theory and ordinal logistic regression. *Children Psychiatry and Human Development, 45*(3), 369-376.

Khan, A., Chien, C.-W., & Brauer, S. G. (2013). Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. *Journal of Clinical Epidemiology, 66*, 681-687.

Kiefer, T., Robitzsch, A., & Wu, M. (2016). Package 'TAM'. Available on http://www.edmeasurementsurveys.com/TAM/Tutorials/

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Kook, S. H., & Varni, J. W. (2008). Validation of the Korean version of the pediatric quality of life inventory™ 4.0 (PedsQL™) generic core scales in school children and adolescents using the Rasch model. *Health & Quality of Life Outcomes, 6*, 41.

Limbers, C. A., Newman, D. A., & Varni, J. W. (2008). Factorial invariance of child self-report across age subgroups: a confirmatory factor analysis of ages 5 to 16 years utilizing the PedsQL 4.0 generic core scales. *Value in Health, 11*(4), 659-668.

Lin, C.-Y., Chang, C.-C., Wu, T.-H., & Wang, J.-D. (2016). Dynamic changes of self-stigma, quality of life, somatic complaints, and depression among people with schizophrenia: a pilot study applying kernel smoothers. *Stigma and Health, 1*(1), 29-43.

Lin, C.-Y., Luh, W.-M., Cheng, C.-P., Yang, A.-L., Su, C.-T., & Ma, H.-I. (2013). Measurement equivalence across child self-reports and parent-proxy reports in the Chinese version of the Pediatric Quality of Life Inventory Version 4.0. *Child Psychiatry & Human Development, 44*(5), 583-590.

Lin, C.-Y., Yang, S.-C., Lai, W.-W., Su, W.-C., & Wang, J.-D. Rasch models suggested the satisfactory psychometric properties of the WHOQOL-BREF among lung cancer patients. *Journal of Health Psychology,* 1359105315603474, first published on September 8, 2015.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Link, B. G. (1987). Understanding labeling effects in the area of mental disorders: an assessment of the effects of expectations of rejection. *American Sociological Review, 52*, 96-112.

Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual Review of Sociology, 27,* 363-385.

Link, B. G., Struening, E. L., Neese-Todd, S., Asmussen, S., & Phelan, J. C. (2001). The consequences of stigma for the self-esteem of people with mental illness. *Psychiatric Services, 52*, 1621-2626.

Link, B. G., Struening, E. L., Neese-Todd, S., Asmussen, S., & Phelan, J. C. (2002). On describing and seeking to change the experience of stigma. *Psychiatric Rehabilitation Skills, 6*, 201-231.

Link, B. G., Struening, E. L., Rahav, M., Phelan, J. C., & Nuttbrock, L. (1997). On stigma and its consequences: evidence from a longitudinal study of men with dual diagnoses of mental illness and substance abuse. *Journal of Health & Social Behavior, 38*, 177-190.

Link, G. G., Yang, L. H., Phelan, J. C., & Collins, P. Y. (2004). Measuring mental illness stigma. *Schizophrenia Bulletin, 30*, 511-541.

Livingston, J. D., & Boyd, J. E. (2010). Correlates and consequences of internalized stigma for people living with mental illness: a systematic review and meta-analysis. *Social Science & Medicine, 71*, 2150-2161.

MacCallum, R. C., Widaman, K. F., Zhang, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-99.

Mak, W. W., & Cheung, R. Y. (2010). Self-stigma among concealable minorities in Hong Kong: conceptualization and unified measurement. *American Journal of Orthopsychiatry, 80,* 267-281.

Mak, W. W., Poon, C. Y., Pun, L. Y., & Cheung, S. F. (2007). Meta-analysis of stigma and mental health. *Social Science & Medicine, 65*, 245-261.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 177-198). Thousand Oaks, CA: Sage.

Pescosolido, B. A., & Martin, J. K. (2015). The stigma complex. *Annual Review of Sociology, 41*, 87-116.

Peterson, D., Barnes, A., & Duncan, C. (2008). *Fighting shadows: self-stigma and mental illness: Whawhai Atu te Whakamā Hihira.* Auckland: Mental Health Foundation of New Zealand.

Petscher, Y., & Schatschneider, C. (2012). Validating scores from new assessments: a comparison of classical test theory and item response theory. In G. Tenenbaum, R. C., Eklud, & A. Kamata (Eds.), *Measurement in sport and exercise psychology* (pp. 41-52). Champaign, IL: Human Kinetics.

Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. Fayers & R. D. Hays (Eds.), *Assessing quality of life in clinical trials: Methods of practice* (2nd ed., pp. 55-73). New York: Oxford University Press.

Ritsher (Boyd), J., & Phelan, J. C. (2004). Internalized stigma predicts erosion of morale among psychiatric outpatients. *Psychiatry Research, 129*, 257-265.

Ritsher (Boyd), J., Otilingma, P. G., & Grajales, M. (2003). Internalized stigma of mental illness: psychometric properties of a new measure. *Psychiatry Research, 121*, 31-49.

Rudick, M. M., Yam, W. H., & Simms, L. J. (2013). Comparing countdown- and IRT-based approaches to computerized adaptive personality testing. *Psychological Assessment, 25*(3), 769-779.

Rüsch, N., Zlati, A., Black, G., & Thornicroft, G. (2014). Does the stigma of mental illness contribute to suicidality? *British Journal of Psychiatry, 205*, 257-259.

Samuel, D. B., Simms, L. J., Clark, L. A., Livesley, W. J., & Widiger, T. A. (2010). An item response theory integration of normal and abnormal personality scales. *Personality Disorders: Theory, Research, and Treatment, 1*(1), 5-21.

Sibitz, I., Amering, M., Unger, A., Seyringer, M. E., Bachmann, A., Schrank, B…Woppmann, A. (2011a). The impact of the social network, stigma and empowerment on the quality of life in patient with schizophrenia. *European Psychiatry, 26*, 28-33.

Sibitz, I., Unger, A., Woppmann, A., Zidek, T., & Amering, M. (2011b). Stigma resistance in patients with schizophrenia. *Schizophrenia Bulletin, 37*, 316-323.

Sick, J. (2010). Assumptions and Assumptions and requirements of Rasch requirements of Rasch measurement. *SHIKEN: JALT Testing & Evaluation SIG Newsletter, 14*(2), 23-29.

Simon, G. E., VonKorff, M., Piccinelli, M., Fullerton, C., & Ormel, J. (1999). An international study of the relation between somatic symptoms and depression. *The New England journal of medicine, 341*(18), 1329-1335.

Stevelink, S. A., Wu, I. C., Voorend, C. G., & van Brakel, W. H. (2012). The psychometric assessment of internalized stigma instruments: a systematic review. *Stigma Research and Action, 2*(2), 100-118.

Su, C.-T., Ng, H.-S., Yang, A.-L., & Lin, C.-Y. (2014). Psychometric evaluation of the Short Form 36 Health Survey (SF-36) and the World Health Organization Quality of Life

Scale Brief Version (WHOQOL-BREF) for patients with schizophrenia. *Psychological Assessment*, *26*(3), 980-989.

Tractenberg, R. E. (2010). Classical and modern measurement theories, patient reports, and clinical outcomes. *Contemporary Clinical Trials, 31*(1), 1-3.

Wahl, O. F. (1999). Mental health consumers' experience of stigma. *Schizophrenia Bulletin, 25*, 467-478.

Ware, J. E., Jr., Kosinski, M., Bjorner, J. B., Bayliss, M. S., Batenhorst, A., Dahlöf, C. G.,…Dowson, A. (2003). Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Quality of Life Research*, *12*, 935-952.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*, 370.

Wu, T.-H., Chang, C.-C., Chen, C.-Y., Wang, J.-D., & Lin, C.-Y. (2015). Further psychometric evaluation of the Self-Stigma Scale-Short: measurement invariance across mental illness and gender. *PLoS One, 10*(2), e0117592.

Yanos, P. T., Roe, D., & Lysaker, P. H. (2010). The impact of illness identity on recovery from severe mental illness. *American Journal of Psychiatric Rehabilitation, 13*(2), 73-93.

Yao, G., Chung, C.-W., Yu, C.-F., & Wang, J.-D. (2002). Development and verification of validity and reliability of the WHOQOL-BREF Taiwan version. *Journal of the Formosan Medical Association*, *101*, 342-351.

Yen, W., & Fitzpatrick, A. R. (2006). *Item response theory*. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 111-153). Westport, CT: Praeger.

Table 1 Participant characteristics and self-stigma scores

| | Mean | SD | *n* | % |
|---|---|---|---|---|
| Age (years) | 43.76 | 11.27 | | |
| Age at onset (years) | 31.88 | 11.81 | | |
| Gender | | | | |
| Male | | | 148 | 42.7 |
| Female | | | 199 | 57.3 |
| Education | | | | |
| ≤ Junior high school | | | 100 | 28.8 |
| Senior high school | | | 146 | 42.1 |
| ≥ College | | | 101 | 28.8 |
| Marital status | | | | |
| Married | | | 138 | 39.8 |
| Single | | | 156 | 45.0 |
| Other | | | 53 | 15.2 |
| Diagnoses | | | | |
| Schizophrenia | | | 161 | 45.1 |
| Depressive disorder | | | 98 | 28.2 |
| Bipolar disorder | | | 43 | 12.4 |
| Other | | | 45 | 13.0 |
| ISMI total score | 2.08 | 0.66 | | |
| ISMI Affect score | 2.11 | 0.74 | | |
| ISMI Cognitive score | 1.98 | 0.63 | | |
| ISMI Behavior score | 2.15 | 0.72 | | |
| SSS-S total score | 2.17 | 0.75 | | |

| | | |
|---|---|---|
| SSS-S Affect score | 2.23 | 0.78 |
| SSS-S Cognitive score | 2.18 | 0.77 |
| SSS-S Behavior score | 2.11 | 0.81 |

**Abbreviations:** SD, standard deviation; ISMI, Internalized Stigma of Mental Illness scale; SSS-S, Self-Stigma Scale-Short.

Note: The ISMI total score includes only three domains: Alienation (Affect), Stereotype Endorsement (Cognitive), and Social Withdrawal (Behavior).

ISMI scores were significantly lower than SSS-S scores, except for the Behavior domain.

Table 2 Pearson correlation between ISMI, SSS, WHOQOL-BREF, and DSSS scores

| | ISMI | | | | SSS-S | | | |
|---|---|---|---|---|---|---|---|---|
| | Total score | Affect | Cognition | Behavior | Total score | Affect | Cognition | Behavior |
| ISMI total score | -- | | | | | | | |
| Affect | .947 | -- | | | | | | |
| Cognition | .945 | .836 | -- | | | | | |
| Behavior | .950 | .852 | .850 | -- | | | | |
| SSS-S total score | .859 | .820 | .793 | .828 | -- | | | |
| Affect | .805 | .767 | .755 | .766 | | -- | | |
| Cognition | .819 | .804 | .752 | .771 | .944 | .862 | -- | |
| Behavior | .821 | .764 | .751 | .819 | .942 | .856 | .815 | -- |
| WHOQOL-BREF score | | | | | | | | |
| Physical | −.590 | −.572 | −.536 | −.569 | −.517 | −.464 | −.522 | −.479 |
| Psychological | −.641 | −.633 | −.569 | −.619 | −.583 | −.550 | −.556 | −.550 |
| Social | −.543 | −.522 | −.474 | −.549 | −.485 | −.447 | −.450 | −.477 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Environmental | −.531 | −.514 | −.488 | −.507 | −.492 | −.454 | −.465 | −.478 |
| DSSS score | | | | | | | | |
| Depression | .559 | .549 | .488 | .553 | .452 | .405 | .424 | .457 |
| Somatic | .437 | .422 | .386 | .436 | .373 | .335 | .355 | .370 |

All *p*-values < .01.

**Abbreviations:** ISMI, Internalized Stigma of Mental Illness scale; SSS-S, Self-Stigma Scale-Short; WHOQOL-BREF, World Health Organization questionnaire on the Quality of Life, Short Form; DSSS, Depression and Somatic Symptoms Scale.

Note: The ISMI total score includes only three domains: Alienation (Affect), Stereotype Endorsement (Cognitive), and Social Withdrawal (Behavior).

Table 3 Results of eight confirmatory factor analysis models

| Model # | $\chi^2$ | *df* | CFI | IFI | TLI | RMSEA | SRMR |
|---|---|---|---|---|---|---|---|
| 1 | 354.27* | 149 | .987 | .987 | .986 | .066 | .036 |
| 2 | 87.85* | 24 | .988 | .988 | .982 | <u>.089</u> | .028 |
| 3 | 1016.53* | 349 | .983 | .983 | .981 | .078 | .043 |
| 4 | 1258.55* | 347 | .976 | .976 | .974 | <u>.100</u> | .045 |
| 5 | 817.30* | 321 | .987 | .987 | .985 | .068 | .033 |
| 6 | 772.04* | 319 | .988 | .988 | .986 | .064 | .034 |
| 7 | 946.72* | 328 | .984 | .984 | .981 | .075 | .042 |
| 8 | 771.43* | 318 | .988 | .988 | .986 | .065 | .035 |

Model 1: first-order construct of Internalized Stigma of Mental Illness scale (ISMI).

Model 2: first-order construct of Self-Stigma Scale-Short (SSS-S).

Model 3: correlated methods (ISMI and SSS-S).

Model 4: correlated traits (Affect, Cognitive, and Behavior).

Model 5: one general trait (Self-stigma) correlated methods (ISMI and SSS-S).

Model 6: correlated traits (Affect, Cognitive, and Behavior) uncorrelated methods (ISMI and SSS-S).

Model 7: correlated traits (Affect, Cognitive, and Behavior) correlated methods minus 1 (ISMI).

Model 8: correlated traits (Affect, Cognitive, and Behavior) correlated methods (ISMI and SSS-S).

CFI = comparative fit index; IFI = incremental fit index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

All fit-indices are acceptable except for two <u>underlined</u> indices, which are close to acceptable.

* $p < .01$.

Correlation coefficients between ISMI and SSS-S ($r$) = .903 (Model 3), .680 (Model 5), and .911 (Model 8).

**Note:** The ISMI includes only three domains: Alienation (Affect), Stereotype Endorsement (Cognitive), and Social Withdrawal (Behavior).

Table 4 Model comparisons

| | Subtrahend; $\Delta\chi^2$ ($\Delta df$)[a] | | | | |
|---|---|---|---|---|---|
| Minuend | Model 4 | Model 7 | Model 5 | Model 6 | Model 8 |
| Model 3 | −242.02 (2)[b] | 69.81 (21)* | 199.23 (28)* | 244.49 (30)* | 245.10 (31)* |
| Model 4 | | 311.83 (19)* | 441.25 (26)* | 486.51 (28)* | 487.12 (29)* |
| Model 7 | | | 129.42 (7)* | 174.68 (9)* | 175.29 (10)* |
| Model 5 | | | | 45.26 (2)* | 45.87 (3)* |
| Model 6 | | | | | 0.61(1) |

Please see model interpretations at the footnotes of Table 3.

[a] $\Delta\chi^2$ and $\Delta df$ are the differences of $\chi^2$ and the *df* between every two models.

[b] Usually the model with more *df* has greater $\chi^2$; however, Model 3 had smaller $\chi^2$ and more *df* than Model 4 did. Therefore, the $\Delta\chi^2$ is minus, and cannot compute for *p*-value.

*$p < .05$

Note: The models are ordered based on their *df* values.

Table 5 Two-parameter logistic item-response-theory analysis of the ISMI and the SSS-S

| **Domain** | | | | Difficulty[a] | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item | Scale | Item Description | Discrimination | 1 | 2 | 3 | 4 | Infit | Outfit |
| **Affect (Alienation)** | | | | | | | | | |
| A1 | ISMI | I feel out of place in the world because I have a mental illness. | 1.88 | −0.90 | −0.58 | 2.10 | 2.90 | 1.02 | 0.98 |
| A2 | ISMI | Having a mental illness has spoiled my life. | 2.42 | −1.40 | −1.02 | 1.69 | 3.32 | 1.04 | 0.91 |
| A3 | ISMI | People without mental illness could not possibly understand me. | 1.47 | −3.90 | −0.95 | 0.87 | 2.55 | 1.01 | 1.07 |
| A4 | ISMI | I am embarrassed or ashamed that I have a mental illness. | 4.07 | −3.29 | −2.29 | 1.27 | 4.57 | 1.02 | 0.88 |
| A5 | ISMI | I am disappointed in myself for having a mental illness. | 3.14 | −5.80 | −1.57 | 1.36 | 3.30 | 1.01 | 0.95 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| A6 | ISMI | I feel inferior to others who don't have a mental illness | 1.96 | −4.97 | −1.33 | 0.14 | 2.61 | 1.01 | 1.05 |
| A7 | SSS-S | I feel that others would know that I have a mental illness. | 1.67 | −2.08 | −1.36 | 0.13 | 2.31 | 1.00 | 1.04 |
| A8 | SSS-S | I feel uncomfortable because I have a mental illness. | 2.90 | −4.03 | −1.82 | 1.46 | 3.70 | 1.03 | 0.90 |
| A9 | SSS-S | I feel like I cannot do anything about my mental illness status. | 1.54 | −1.14 | −0.87 | 1.88 | 2.41 | 1.07 | 1.01 |
| **Cognitive (Stereotype endorsement)** | | | | | | | | | |
| C1 | ISMI | Stereotypes about people with mental illness apply to me. | 0.93 | −4.79 | −0.43 | 1.28 | 2.55 | 1.02 | 1.04 |
| C2 | ISMI | People can tell that I have a | 1.12 | −2.56 | −0.59 | 1.02 | 1.86 | 1.01 | 1.05 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | mental illness by the way I look. | | | | | | | |
| C3 | ISMI | People with mental illness tend to be violent. | 1.79 | −3.76 | −0.75 | 1.34 | 2.76 | 1.01 | 1.00 |
| C4 | ISMI | Because I have a mental illness, I need others to make most decisions for me. | 2.49 | −3.14 | −1.10 | 1.87 | 4.66 | 1.03 | 0.93 |
| C5 | ISMI | People with mental illness cannot live a good, rewarding life. | 1.78 | −2.89 | −0.85 | 1.60 | 3.31 | 1.04 | 0.99 |
| C6 | ISMI | People with mental illness shouldn't get married. | 2.18 | −3.88 | −0.99 | 2.22 | 3.81 | 1.03 | 0.90 |
| C7 | ISMI | I can't contribute anything to society because I have a mental | 1.62 | −2.66 | −0.92 | 1.54 | 2.45 | 1.01 | 1.07 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | illness. | | | | | | | |
| C8 | SSS-S | My identity as having mental illness incurs inconvenience in my daily life. | 2.24 | −4.72 | −1.65 | 0.67 | 3.29 | 1.02 | 0.99 |
| C9 | SSS-S | My identity having mental illness is a burden to me. | 2.58 | −5.01 | −1.96 | 0.42 | 3.44 | 1.02 | 0.96 |
| C10 | SSS-S | The identity of having mental illness taints my life. | 3.09 | −3.13 | −1.48 | 1.83 | 3.98 | 1.01 | 0.91 |
| **Behavior (Social withdrawal)** | | | | | | | | | |
| B1 | ISMI | I don’t talk about myself much because I don’t want to burden others with my mental illness. | 1.86 | −4.49 | −0.78 | 2.00 | 3.60 | 1.05 | 0.94 |
| B2 | ISMI | I don’t socialize as much as I used | 2.28 | −3.82 | −1.12 | 0.81 | 3.17 | 1.02 | 0.99 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | to because my mental illness might make me look "weird" or behave strangely. | | | | | | | |
| B3 | ISMI | Negative stereotypes about mental illness keep me isolated from the "normal" world. | 1.18 | −1.43 | −1.18 | −0.72 | 1.68 | 1.02 | 1.01 |
| B4 | ISMI | I stay away from social situations in order to protect my family or friends from embarrassment. | 2.32 | −2.28 | −1.61 | 1.25 | 3.15 | 1.03 | 0.99 |
| B5 | ISMI | Being around people who don't have a mental illness makes me feel | 2.44 | −4.60 | −1.21 | 1.44 | 3.64 | 1.04 | 0.91 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | out of place or inadequate. | | | | | | | |
| B6 | ISMI | I avoid getting close to people who don’t have a mental illness to avoid rejection. | 2.98 | −3.30 | −1.73 | 1.52 | 3.94 | 1.02 | 0.92 |
| B7 | SSS-S | I dare not to make new friends lest they find out that I have a mental illness. | 3.04 | −2.76 | −1.93 | 1.79 | 4.34 | 1.01 | 0.97 |
| B8 | SSS-S | I estrange myself from others because I have a mental illness. | 5.37 | −4.89 | −3.17 | 2.48 | 6.42 | 1.02 | 0.78 |
| B9 | SSS-S | I avoid interacting with others because I have a mental illness. | 3.90 | −3.82 | −1.89 | 1.94 | 5.23 | 1.03 | 0.83 |

Abbreviations: ISMI, Internalized Stigma of Mental Illness scale; SSS-S, Self-Stigma Scale-Short; Infit, information-weighted fit statistics; Outfit, outlier-sensitive fit statics.

[a] Difficulty for each category

# Appendices

**Appendix A: Results using domains of Discrimination Experience and Stigma Resistance in the Internalized Stigma Mental Illness scale (ISMI)**

**I. Concurrent validity**

Table S1Pearson correlation between Discrimination Experience, Stigma Resistance, WHOQOL-BREF, and DSSS scores

| | ISMI total score | Discrimination Experience | Stigma Resistance |
|---|---|---|---|
| WHOQOL-BREF score | | | |
| Physical | −.555* | −.500* | −.151 |
| Psychological | −.593* | −.549* | −.093 |
| Social | −.523* | −.530* | −.128 |
| Environmental | −.493* | −.467* | −.061 |
| DSSS score | | | |
| Depression | .434* | .383* | .020 |
| Somatic | .540* | .464* | .019 |

**Abbreviations:** WHOQOL-BREF, World Health Organization questionnaire on the Quality of Life, Short Form; DSSS, Depression and Somatic Symptoms Scale.
Note: The ISMI total score includes all five domains: Alienation, Stereotype Endorsement, Social Withdrawal, Discrimination Experience, and Stigma Resistance.
* p<.01

## II. Comparisons of confirmatory factor analysis (CFA) models

Table S2 Results of eight confirmatory factor analysis models

| Model # | $\chi^2$ | *df* | CFI | IFI | TLI | RMSEA | SRMR |
|---|---|---|---|---|---|---|---|
| 1 | 933.35* | 367 | .908 | .908 | <u>.898</u> | .067 | .073 |
| 2 | 87.85* | 24 | .988 | .988 | .982 | <u>.089</u> | .028 |
| 3 | 2023.58* | 664 | <u>.860</u> | <u>.861</u> | <u>.852</u> | .077 | .056 |
| 4 | 2051.85* | 655 | <u>.856</u> | <u>.857</u> | <u>.846</u> | .078 | .068 |
| 5 | 1627.49* | 626 | <u>.897</u> | <u>.898</u> | <u>.884</u> | .068 | .038 |
| 6 | 1495.05* | 617 | .910 | .910 | <u>.897</u> | .064 | .061 |
| 7 | 1600.35* | 626 | .900 | .901 | <u>.887</u> | .067 | .062 |
| 8 | 1370.15* | 616 | .922 | .923 | .911 | .059 | .039 |

Model 1: first-order construct of Internalized Stigma of Mental Illness scale (ISMI).
Model 2: first-order construct of Self-Stigma Scale-Short (SSS-S).
Model 3: correlated methods (ISMI and SSS-S).
Model 4: correlated traits (Affect, Cognitive, Behavior, Discrimination Experience, and Stigma Resistance).
Model 5: one general trait (Self-stigma) correlated methods (ISMI and SSS-S).
Model 6: correlated traits (Affect, Cognitive, Behavior, Discrimination Experience, and Stigma Resistance) uncorrelated methods (ISMI and SSS-S).
Model 7: correlated traits (Affect, Cognitive, Behavior, Discrimination Experience, and Stigma Resistance) correlated methods minus 1 (ISMI).
Model 8: correlated traits (Affect, Cognitive, Behavior, Discrimination Experience, and Stigma Resistance) correlated methods (ISMI and SSS-S).
CFI = comparative fit index; IFI = incremental fit index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. Unsatisfactory fit indices are <u>underlined</u>.
* $p < .01$.
Note: The ISMI includes all five domains: Alienation (Affect), Stereotype Endorsement (Cognitive), and Social Withdrawal (Behavior), Discrimination Experience, and Stigma Resistance.

Table S3 Model comparisons for CFAs

| | Subtrahend; $\Delta\chi^2$ ($\Delta df$)[a] | | | | |
|---|---|---|---|---|---|
| Minuend | Model 4 | Model 5 | Model 7 | Model 6 | Model 8 |
| Model 3 | −28.27 (9)[b] | 396.09 (38)* | 423.23 (38)* | 528.53 (47)* | 653.44 (48)* |
| Model 4 | | 424.36 (29)* | 451.50 (29)* | 556.80 (38)* | 681.71 (39)* |
| Model 5 | | | 27.14 (0)[c] | 132.44 (9)* | 257.35 (10)* |
| Model 7 | | | | 105.30 (9)* | 230.20 (10)* |
| Model 6 | | | | | 124.91 (1)* |

Please see model interpretations at the footnotes of Table 2.
[a] $\Delta\chi^2$ and $\Delta df$ are the differences of $\chi^2$ and the *df* between every two models.
[b] Usually the model with more *df* has greater $\chi^2$; however, Model 3 had smaller $\chi^2$ and more *df* than Model 4 did. Therefore, the $\Delta\chi^2$ is minus, and cannot compute for *p*-value.
[c] Models 5 and7 had the same *df*, and cannot compute for *p*-value; however, Model 7 outperformed Model 5 because of the smaller $\chi^2$.
* $p <.05$
Note: The models are ordered based on their *df* values.

**III. Tests of unidimensionality and the comparisons of Rasch and 2PL models**

Table S4 Tests of unidimensionality

| **Domain** | Goodness-of-fit index, GFI | Root-mean-square residual, RMSR |
|---|---|---|
| Discrimination Experience | .992 | .012 |
| Stigma Resistance | .975 | .034 |

Table S5 The comparisons of Rasch and 2PL models for Discrimination Experience and Stigma Resistance

| **Domain**<br>Model | AIC | BIC | $\chi^2$ | $\Delta\chi^2$ (*df*) | p |
|---|---|---|---|---|---|
| **Discrimination Experience** | | | | | |
| Rasch model | 3631.470 | 3712.306 | 3589.470 | 47.808 (4) | <.001 |
| 2PL model | 3591.663 | 3687.896 | 3541.663 | | |
| **Stigma Resistance** | | | | | |
| Rasch model | 3879.809 | 3960.644 | 3837.809 | 84.939 (4) | <.001 |
| 2PL model | 3802.870 | 3899.103 | 3752.870 | | |

AIC= Akaike information criterion; BIC= Bayesian information criterion; 2PL model=Two-parameter logistic item-response-theory model

**IV. Results of 2PL models**

Table S6 Two-parameter logistic item-response-theory analysis of Discrimination Experience and Stigma Resistance

| Domain | | | | Difficulty[a] | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item | Scale | Item Description | Discrimination | 1 | 2 | 3 | 4 | Infit | Outfit |
| **Discrimination Experience** | | | | | | | | | |
| D1 | ISMI | People discriminate against me because I have a mental illness. | 1.87 | −2.21 | −1.05 | 1.00 | 3.27 | 1.02 | 1.00 |
| D2 | ISMI | Others think that I can't achieve much in life because I have a mental illness. | 1.15 | −2.10 | −0.67 | 1.23 | 2.24 | 1.04 | 1.07 |
| D3 | ISMI | People ignore me or take me less seriously just because I have a mental illness. | 3.41 | −4.10 | −2.07 | 1.84 | 4.92 | 1.02 | 0.90 |
| D4 | ISMI | People often patronize me, or treat me like a child, just because I have a mental illness. | 2.89 | −2.95 | −1.16 | 2.62 | 4.73 | 1.03 | 0.86 |
| D5 | ISMI | Nobody would be interested in getting close to me because I have a mental illness. | 1.90 | −4.17 | −1.36 | 0.84 | 3.07 | 1.02 | 1.00 |
| **Stigma Resistance** | | | | | | | | | |
| S1 | ISMI | I feel comfortable being seen in public with a person who obviously has a mental illness. | 0.68 | −1.74 | −1.06 | −1.41 | 1.21 | 1.01 | 1.01 |
| S2 | ISMI | In general, I am able to live life the way I want to. | 0.26 | −2.12 | −0.64 | 0.39 | 1.60 | 1.00 | 1.00 |
| S3 | ISMI | I can have a good, fulfilling life, despite my mental illness. | 0.77 | −3.20 | −1.42 | −0.56 | 1.71 | 1.01 | 1.03 |
| S4 | ISMI | People with mental illness make important contributions to society. | 2.24 | −4.89 | −3.19 | −2.02 | 2.71 | 1.01 | 1.00 |
| S5 | ISMI | Living with mental illness has made me a tough survivor. | 2.63 | −4.13 | −3.89 | −2.62 | 2.62 | 0.99 | 0.94 |

Abbreviations: ISMI, Internalized Stigma of Mental Illness scale; Infit, information-weighted fit statistics; Outfit, outlier-sensitive fit statics.

[a] Difficulty for each category